

Development of a Gaussian Process – feature selection model to characterise (poly)dimethylsiloxane (Silastic®) membrane permeation

Yi Sun^a, Mark Hewitt^b, Simon C. Wilkinson^c, Neil Davey^a, Roderick G. Adams^a, Darren R. Gullick^{d,*} and Gary P. Moss^e 

^aSchool of Computer Science, University of Hertfordshire, Hatfield, ^bSchool of Pharmacy, University of Wolverhampton, Wolverhampton, ^cSchool of Biomedical, Nutritional and Sports Sciences, Medical School, University of Newcastle-upon-Tyne, Newcastle-upon-Tyne, ^dSchool of Pharmacy & Biomedical Sciences, University of Portsmouth, Portsmouth, ^eThe School of Pharmacy, Keele University, Keele, UK

Keywords

data set design; feature selection; Gaussian Process Regression; machine learning; polydimethylsiloxane

Correspondence

Gary P. Moss, The School of Pharmacy, Keele University, Keele, Staffordshire ST5 5BG, UK.
Email: g.p.j.moss@keele.ac.uk

Received January 17, 2020

Accepted March 8, 2020

doi: 10.1111/jphp.13263

*Present address: ETS Laboratories, Saint Helena, CA, 94574, USA

Abstract

Objectives The current study aims to determine the effect of physicochemical descriptor selection on models of polydimethylsiloxane permeation.

Methods A total of 2942 descriptors were calculated for a data set of 77 chemicals. Data were processed to remove redundancy, single values, imbalanced and highly correlated data, yielding 1363 relevant descriptors. For four independent test sets, feature selection methods were applied and modelled via a variety of Machine Learning methods.

Key findings Two sets of molecular descriptors which can provide improved predictions, compared to existing models, have been identified. Best permeation predictions were found with Gaussian Process methods. The molecular descriptors describe lipophilicity, partial charge and hydrogen bonding as key determinants of PDMS permeation.

Conclusions This study highlights important considerations in the development of relevant models and in the construction and use of the data sets used in such studies, particularly that highly correlated descriptors should be removed from data sets. Predictive models are improved by the methodology adopted in this study, notably the systematic evaluation of descriptors, rather than simply using any and all available descriptors, often based empirically on *in vitro* experiments. Such findings also have clear relevance to a number of other fields.

Introduction

Predicting the permeation into and across the skin of exogenous chemical is of substantial interest for a number of industries, including pharmaceuticals, cosmetics, pesticides and the handling of industrial chemicals. Determining skin permeability is therefore an essential component of understanding the risks associated with exposure of the skin to exogenous chemicals. In general, *in vitro* experiments form a significant part of early-stage evaluation of pharmaceutical formulations or in risk assessment protocols for other topical exposures. While fresh human skin is the perceived 'gold standard' for *in vitro* testing, its consistent use is constrained by availability, which often means that certain compromises are commonly adopted,

including the use of previously frozen human skin and tissue from other species; in the latter case, it is generally accepted that pig skin is the best model for human skin, with the pig ear being widely used despite differences in the lateral packing of *stratum corneum* lipids, compared to human skin.^[1-4]

Given the scientific and logistical constraints discussed above, artificial membranes have also found widespread use in early-stage assessment of percutaneous absorption, notably in the context of formulation optimisation and elucidation of permeation mechanisms. Polydimethylsiloxane (PDMS, Figure 1) is one of the most widely used membranes in such studies and has been shown to correlate well with mammalian skin studies^[5] but differences in the distribution of permeation compared to mammalian skin have

also been found.^[6,7] Similarly, Gullick and co-workers found reasonable correlations between *in vitro* diffusion experiments using PDMS membranes and pig skin.^[8] PDMS has also been widely used to investigate mechanisms of membrane transport. Permeability was generally related to the physicochemical properties of their penetrants, and solvents were taken up into the membrane, altering membrane properties and the flux of the permeants.^[9–12] ATR-FTIR spectroscopy has also been used to evaluate diffusion across a PDMS membrane^[13] with a rudimentary structure–activity relationship for permeability across a PDMS membrane being developed.^[14] It was also reported that the permeation distribution of a range of chemicals across a PDMS membrane was Gaussian-normal,^[15] in contrast to a number of studies which reported non-Gaussian (log-normal) distribution in mammalian skin.^[6,9,16–21] This difference was broadly attributed to the heterogeneity of biological membranes, including the possibility of multiple permeation pathways in mammalian skin, which is in contrast to the homogeneity of PDMS membranes. PDMS membranes are therefore pharmaceutically significant and provide an effective screen in early-stage formulation development and in the elucidation of mechanistic information for the permeation process.

The first major studies quantifying permeability across a PDMS membrane initially saw the development of empirical models for permeation across a PDMS membrane which related flux through the membrane to partial atomic charge, mole fraction solubility and molecular weight.^[22]

$$\log J_{\text{mss}} = -2.497 - 4.339\Sigma e_+ - 1.531\Sigma e_- + 4.065(\Sigma e_+ \cdot \Sigma e_p) + 0.649 \log C_s - 0.00651MW - 0.640\text{imidazole} + 0.689\text{amine} \quad (1)$$

$n = 103$, $r^2 = 0.966$, $s = 0.238$, $F = 386.5$; where J_{mss} is the maximum steady-state flux ($\mu\text{mol/s per cm}^2$); Σe_+ is the sum of the charge values of hydrogen atoms (with charge >0.1) and the positive charge of a nitrogen atom in a nitro group; Σe_- is the sum of the absolute charge values of all other heteroatoms with unshared electron pairs in the same molecule.

Equation 1 gave better predictions than their previous model, and they applied it to predict the flux of 171 new

compounds which were not included in their previous study.^[23] This yielded a slightly simplified model which omitted the imidazole descriptor:

$$\log J_{\text{mss}} = -2.497 - 4.339\Sigma e_+ - 1.531\Sigma e_- + 4.065(\Sigma e_+ \cdot \Sigma e_-) + 0.649 \log C_s - 0.00651MW + 0.689\text{amine} \quad (2)$$

These data^[22,23] were re-analysed with the aim of developing a QSAR model based on readily calculable descriptors, unlike those used in the original studies, with greater mechanistic insight for the whole data set.^[24] This resulted in a simple QSAR:

$$\log J = -0.561\text{HA} - 0.671\text{HD} - 0.801^6\chi - 0.383 \quad (3)$$

$n = 242$, $r = 0.900$, $s = 0.464$, $F = 338$; where J is the steady-state flux ($\mu\text{mol/s per cm}^2$), HA and HD are, respectively, the number of hydrogen bond acceptor and donor groups present on a penetrant and $^6\chi$ is the sixth-order path molecular connectivity.

This model describes permeability across a PDMS membrane in terms of hydrogen bonding and molecular topology. These data^[22–24] were also modelled by in an artificial neural network (ANN) study.^[25] They generated a 12-parameter non-linear QSAR model which, most significantly, found that $\log P$ was not significant, being attributed to the inability of $\log P$ to account for intramolecular interactions. Similar findings were also reported^[26] although these were based on a very small data set ($n = 16$). In more recent years, Gaussian Processes Regression methods (GP, or GPR) have been shown to outperform QSARs in predicting the skin permeability of compounds.^[27] GPs have recently been used for the prediction of the permeability of compounds across non-human skin and synthetic polydimethylsiloxane membranes,^[28,29] and human skin.^[30]

Two of the major problems with developing valid and useful models of biological functions, such as skin permeability, are the selection of appropriate descriptors (features) and the size of the data set. The impact for a small data set may be more prominent as the solution (the set of selected descriptors) may change more significantly if one data point is replaced by another. These issues have been addressed in recently,^[30–32] resulting in the models that are highly predictive and robust.

It is clear that a number of previous studies in this field focus on using existing data or collating as much available data as possible and that the descriptors of interest are often chosen based on empirical studies. Such approaches limit the value of models, and while a detailed understanding of data set size and distribution has been addressed previously,^[32] this study focuses on the selection of objective relevant physicochemical descriptors, employing a Feature

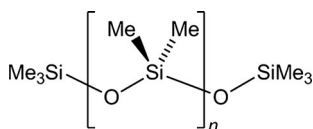


Figure 1 Chemical structure of (poly)dimethylsiloxane (PDMS).

Selection methodology to do so. Therefore, the two aims of this work were, firstly, to find a set of highly relevant molecular descriptors that describe the process of membrane permeation and, secondly, to determine whether feature selection techniques can yield benefits in the prediction of chemical transport across PDMS membranes and to provide a more specific and nuanced understanding of the mechanistic nature of the permeation process. The novelty of this work is that this is the first time such methodology (that is, feature selection on a small data set with a large number of physicochemical descriptors) has been applied to a system of pharmaceutical interest and this clearly has implications for other permeation across other membranes, notably skin.

Methods

Description of the data set

A human skin data set was not used in this study due to previously reported inconsistencies in the available data,^[33–35] which it was felt could inhibit the development of new models based on the approach used in this study. Thus, a PDMS membrane was considered more appropriate as its use is generally associated with a greater consistency.^[22–24,36] The data set was collated from the existing scientific literature and is available as Appendix S2. It consists of 77 unique chemical compounds. For each compound, the permeability coefficient, k_p (cm/h), across a PDMS membrane was used and 2942 physicochemical descriptors were determined using a range of software packages (Dragon Professional, Molecular Operating Environment (MOE), HYdrogen BOnd Thermodynamics (HYBOT) and WinMolconn). While it is more common to use flux (J) to describe membrane permeation,^[22–24] the permeability coefficient, k_p , which is effectively the concentration-corrected flux, is more widely used in the scientific literature and, notably, in the construction of skin permeation models. This is because the use of k_p allows a comparison to be made between different chemicals as concentration differences are generally accounted for.^[36–38]

Data set pre-processing

There are no missing values or inconsistent data in the data set used in this study. 772 descriptors had only one possible value and were removed from the data set. A further 194 descriptors were removed as their distributions were highly imbalanced (i.e., the number of one certain value is greater than a threshold set in this analysis, i.e. 90% or greater of the total number of data points). 613 further descriptors which were highly correlated within the data set (i.e. those whose correlation coefficients are greater than 0.99) were removed. The analysis in this study was conducted on a

refined data set of 77 compounds, each with 1363 physicochemical descriptors.

Generation of four independent test sets

It has been demonstrated previously that there are no significant linear relationships with the target permeability data for the descriptors used.^[37,39] Principal component analysis (PCA) indicated that no linear relationship was observed between the permeability coefficient and the compound descriptors combined in the first component, where 40.2% of the variance was due to the first two principal components, PC1 and PC2. Figure 2 shows a PCA plot of the original, complete, PDMS data set, indicating that the first two Eigenvectors capture 40.21% of the total data variance. In order to observe how the final selected descriptor set affects the performance of regression models on the test set, four different test sets were randomly constructed, each of which includes seven test compounds:

- Set A – seven test compounds were selected from the boundary of the data cloud shown in the PCA (Figure 3a for PC1, and Figure 3b for PC2, denoted by crosses, ‘+’ in these figures). These PCA plots further show the relationship between $\log K_p$ and the first two principal components, indicating that, except for the selected compound with the largest $\log K_p$ value and the two selected compounds with relatively small $\log K_p$ values, the remaining four selected compounds, compounds with different PC projections but very similar $\log K_p$ values can be found. This test set is denoted *TstA* and the rest of the chemical compounds as *TrnA*.
- Set B – this test set includes the three smallest and the four largest $\log K_p$ values. Figure 3c,d shows projections of these compounds on a PCA plot for PC1 and PC2, respectively; this test set is denoted as *TstB* and the rest of the chemical compounds as *TrnB*.
- Set C – seven compounds randomly selected from the data set. Figure 4a,b shows the projections of these compounds in the PCA plot for PC1 and PC2, respectively. This test set is denoted as *TstC* and the remaining chemical compounds as *TrnC*.
- Set D – seven compounds selected from the t-sne plot^[40] and for PC1 and P2 is shown in Figure 4c,d, respectively. This is denoted as *TstD* and the rest of the chemicals as *TrnD*.

Constructing random training/validation sets

To obtain robust results for each training set the data has been shuffled 10 000 times. Each shuffle sees 60 chemicals randomly selected as the training set of regression models and the remaining compounds are used as the validation

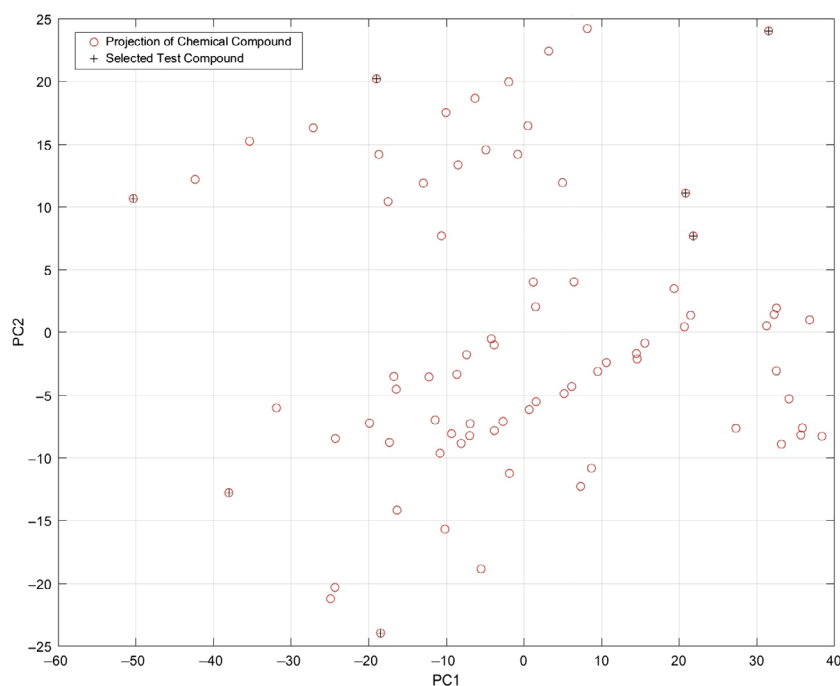


Figure 2 A principal component analysis plot of the original PDMS data set, showing that the first two Eigenvectors capture 40.21% of the total variance. [Colour figure can be viewed at wileyonlinelibrary.com]

set.^[28-32,37,39] Thus, there are 10 000 training subsets and 10 000 validation sets corresponding to each training set. The validation sets are used for feature selection rather than hyperparameter tuning. In using such a method, descriptors selected from different training sets will vary and those that are consistently selected are of most interest. The standard deviation is also used to describe the variance of the outputs.

Methods of analysis

In this study, a range of methods were used. These methods are outlined below; they have been reported on extensively previously^[30-32,37,39,41-59] and detailed methods are also included as Appendix S1.

Random Forest Trees (RFTs)

The Random Forests algorithm was coupled with the random selection of descriptors and bootstrap aggregation to the training sets.^[52] In this study, an ensemble of decision trees for regression has been applied via MatLab *TreeBagger*. The 50 most important descriptors in each ensemble of decision trees for a given training set were identified. Thus, either a trained RFT or GP model has been developed using the most relevant physicochemical descriptors only. While normally it would not be necessary to have a validation set

for the RFT experiments, we have, in this case, used a validation set in order to keep our results consistent with the other tests used in this study.

LASSO

In this study, the LASSO method^[54] has been applied, with a five-fold cross-validation method, in order to remove redundant descriptors from the data set. To speed up the processing time the *UseParallel* mode has been enabled. All other parameters are set to the default mode in MATLAB.

(Joint) Mutual Information (MI, JMI)

In this study, the MatLab FEAST toolbox (<http://www.cs.man.ac.uk/gbrown/fstoolbox>) has been used to select descriptors using the JMI criterion by assigning labels to the responses required by the model. Those responses have been aligned into 10 categories.

Gaussian Processes

A Gaussian Process (GP) is defined as a collection of random variables which, jointly, have a Gaussian distribution.^[57] GPs have also been successfully applied to the field of predicting percutaneous absorption.^[30-32,37] GP methods have been described in detail previously and are also discussed in the Appendix S1.

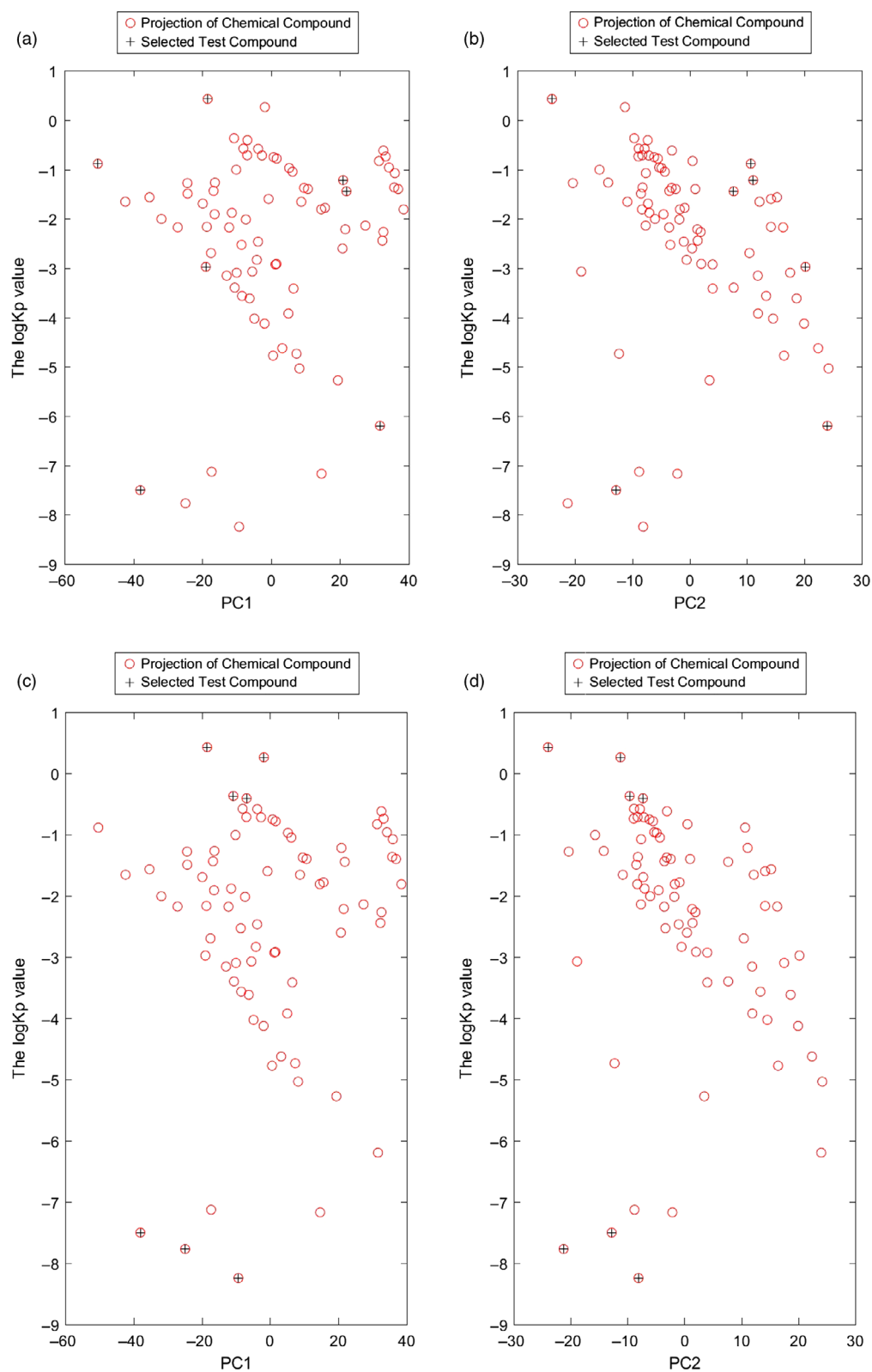


Figure 3 Boundary points in the principal component analysis plot shown for (a) PC1 and (b) PC2 for Test Set A, and with the three smallest logK_p values and four largest logK_p values highlighted for (c) PC1 and (d) PC2 for Test Set B. [Colour figure can be viewed at wileyonlinelibrary.com]

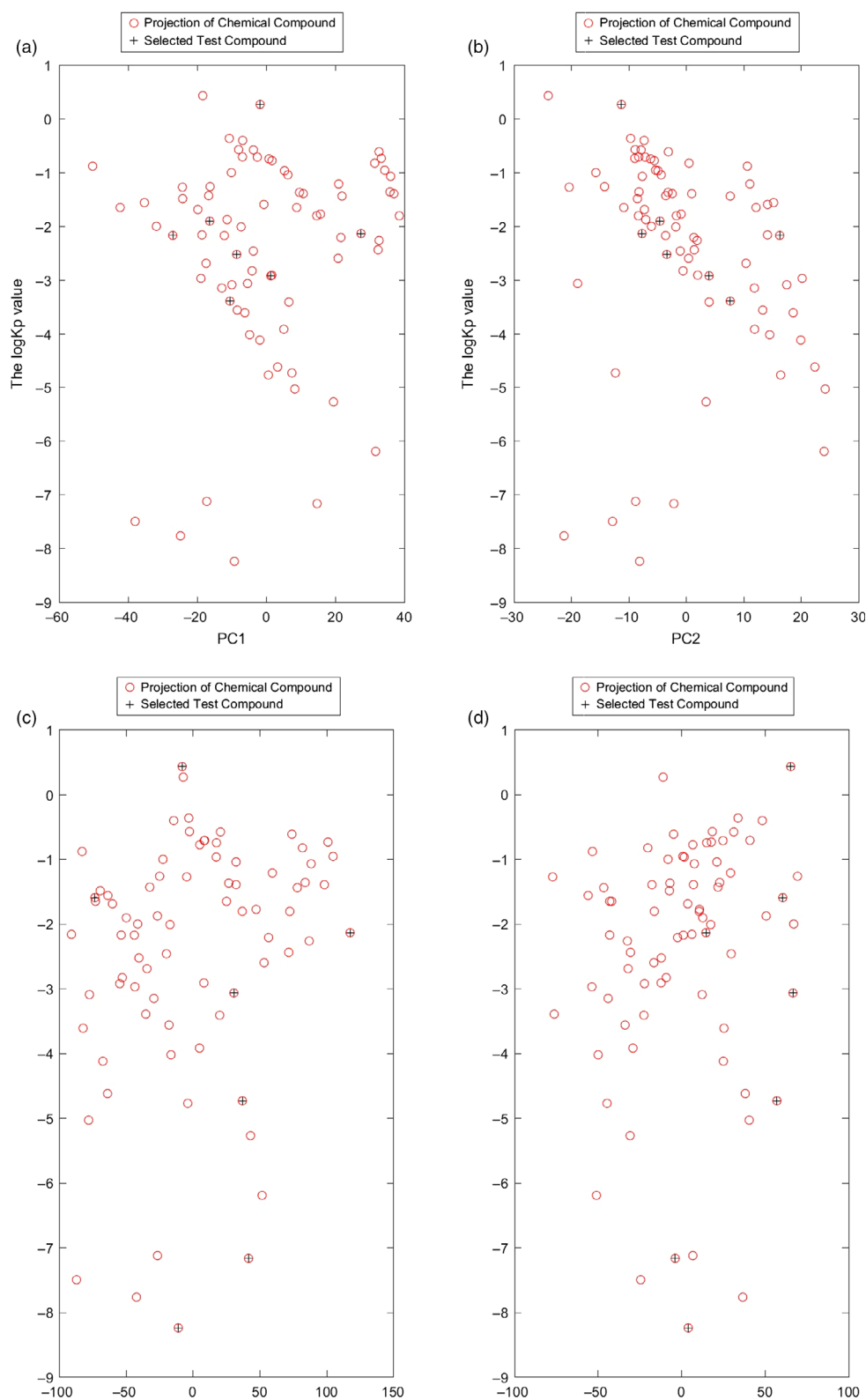


Figure 4 Principal component analysis plots for Test Sets C, showing seven compounds randomly selected from the data set for (a) PC1 and (b) PC2, and for Test Set D showing seven compounds selected from the t-sne plot,^[40] for (c) PC1 and (d) PC2. [Colour figure can be viewed at wileyonlinelibrary.com]

Automatic Relevance Determination

To implement automatic relevance determination (ARD) in a GPR, the characteristic length scale matrix, M , is redefined as a diagonal matrix containing the elements of vector $L = [l_1^{-2}, \dots, l_D^{-2}]$ and l_1, \dots, l_D on the diagonal are the characteristic length scales for each input dimension, determining how relevant an input is to the task. These characteristic length scales can be optimised from the data by Bayesian inference.

Performance measures

The mean squared error (MSE), improvement over Naïve model (ION) and the Pearson correlation coefficient (r , or CORR) we are all used to evaluate the performance of each model.^[30–32,37,39,59] MSE measures the average squared difference between model predictions and the corresponding targets. The ION measures the degree of improvement of the model performance over the performance of the Naïve predictor, which is normally the mean of the output (e.g. $\log K_p$) in the training set. The ION is thus defined as:

$$\text{ION} = \frac{\text{MSE}_{\text{Naïve}} - \text{MSE}}{\text{MSE}_{\text{Naïve}}} \times 100\% \quad (4)$$

CORR measures the correlation between predictions and targets. For comparison, a ‘good’ model should have a low value of MSE and high values of ION and CORR on a given test data set.

Results and Discussion

In this field, it is important to note that previous studies are either based on small data sets (e.g. 16 chemicals of limited structural diversity^[26]) or on different experimental conditions (e.g. an isopropylalcohol solvent system at 30°C^[22–24]). Benchmarking is therefore a challenge, and the decision to focus on a skin permeation QSAR^[60] is limited, but offers a relevance in terms of comparison of PDMS relevance to skin.^[39] A consideration of the ION values between the models and the benchmark suggests a limited ability of data from PDMS membranes to successfully model skin, and *vice versa*, emphasising the limitations of the PDMS membrane in acting as a substitute for skin. Table 1 also shows the results of applying GP models.^[30,31] to these data and does so using the same parameters as the QSAR model in order to provide a comparable benchmark as well as a full set of 1363 descriptors. While it is usual for two-thirds of the data set to be used for training and the remaining one-third for the test in this work, in aiming to train the model with as many data points as possible and to repeat this 10 000 times to address concerns of inherent

variance, we aimed for 10% test data, keeping most data in the training set so the estimates of permeability were as accurate as possible. We were concerned that if we removed more data from the training set the predictions would be significantly impaired, a finding which was apparent in previous work with similar types of data.^[32,61,62]

We have previously used several measures of model quality, including MSE and ION, rather than just the correlation coefficient.^[28–30,37] This is important because in this study (e.g. Table 1) it is clear that the QSAR models have the lowest ION values but reasonably high correlation coefficients, suggesting a possible systematic error. The GP models have comparably improved ION values. It is interesting to note that the MSE and ION results are better than for the QSAR model but still very poor for the correlation coefficients. This might reflect the somewhat comparable nature of the linear QSAR model with the experimental findings that associate permeation across a PDMS membrane to be predominately linear – thereby over-estimating permeability particularly at high lipophilicities. This is not reflected in the GP models which are more reflective of the non-linear nature of skin permeability in the context of molecular descriptors such as $\log P$ and molecular weight.^[37] It also highlights the limitations of replacement membranes – in this case, PDMS and mirrors similar outcomes reported previously for a range of mammalian species.^[39]

It should also be noted that the results for the GP models shown in Table 1 do indicate that ION and MSE values are significantly better for these models, even if the variance associated with each outcome is large when the full set of descriptors is used. This suggests that the full set of descriptors is responsible for noise and redundancy within the data, highlighting the importance of selecting the correct descriptors as well as the need to rationalise and optimise the descriptors used for our models in order to produce more accurate estimations of membrane permeability. Further, the ION values are comparable to those for the correlation coefficient and provide similar measurements of model quality. This is important as it allows readers from different disciplines to frame the outcomes of this study in their own contexts.

Table 2 summarises the results of the first set of experiments on the PDMS data set. These results indicate that the best overall model is found using the Random Forests method, which outperforms other models for three of the four validation sets – the Mutual Information method performs better on validation set D. This is most likely because the Random Forests method embeds feature selection methodology in regression. It was also found that the results of the LASSO method were not stable and that the mean values of estimations on two validation sets – A and C – were worse than the naïve predictions. Further, when the GPARD + Full method is compared to GP it was observed that the use

Table 1 Results on data sets with two features (MW and log P) using (a) the benchmark QSAR model of Potts and Guy for skin^[60]; (b) the Gaussian Process Regression model with two descriptors (MW and log P) and (c) the Gaussian Process Regression model with the full set of descriptors. Results are the mean value and standard deviation over 10 000 validation sets, which are denoted as *ValA*, *ValB*, *ValC* and *ValD*, respectively

	ValA	ValB	ValC	ValD
QSAR model				
MSE	15.98 ± 1.61	16.23 ± 1.89	16.07 ± 2.22	16.31 ± 1.70
ION%	−627.35 ± 507.95	−871.35 ± 555.31	−544.81 ± 405.01	−683.68 ± 527.58
CORR	0.55 ± 0.32	0.66 ± 0.25	0.57 ± 0.29	0.62 ± 0.35
Gaussian Process model with two descriptors				
MSE	3.13 ± 2.32	1.76 ± 1.61	3.84 ± 2.58	2.64 ± 1.78
ION%	6.20 ± 62.70	22.36 ± 85.27	3.73 ± 53.42	9.76 ± 45.79
CORR	0.37 ± 0.46	0.59 ± 0.35	0.34 ± 0.41	0.39 ± 0.31
Gaussian Process model with full descriptors				
MSE	3.25 ± 1.83	2.32 ± 1.27	3.56 ± 2.01	2.95 ± 1.73
ION%	−0.00 ± 0.03	−2.04 ± 9.01	−0.02 ± 0.10	0.08 ± 0.74
CORR	0.12 ± 0.43	0.21 ± 0.37	0.21 ± 0.30	0.36 ± 0.13

QSAR is quantitative structure–activity relationship, and the term QSPR (quantitative structure–permeability) relationship is often interchangeably used in modelling studies of skin permeation; MSE is mean squared error; ION, or ION%, is the improvement over the naïve model; CORR, or *r*, is the correlation coefficient; MW is molecular weight; log P is the octanol : water partition coefficient. Results highlighted in bold are those which achieved the best outcome for each test.

Table 2 Summary of the statistical measures for the validation sets used to assess model quality for the PDMS data set. The results are presented as (a) MSE; (b) ION and (c) the correlation coefficient

	ValA	ValB	ValC	ValD
MSE				
RFTs + RFTs	2.46 ± 2.06	1.49 ± 1.50	2.84 ± 2.12	1.82 ± 1.93
GP + RFTs	2.86 ± 1.99	2.22 ± 1.43	3.43 ± 2.22	1.65 ± 1.80
Lasso + Lasso	5.70 ± 223.31	1.64 ± 1.64	3.85 ± 2.50	2.33 ± 2.05
GP + Lasso	3.65 ± 8.03	2.06 ± 1.74	3.90 ± 2.84	2.33 ± 2.27
GP + MI	2.56 ± 2.01	2.51 ± 1.33	3.56 ± 2.08	1.42 ± 1.54
GPARD + Full	2.68 ± 1.96	1.98 ± 1.64	3.38 ± 2.16	2.30 ± 1.64
GP + two fs	3.25 ± 1.83	2.32 ± 1.27	3.56 ± 2.01	2.95 ± 1.73
ION				
RFTs + RFTs	30.17 ± 33.50	44.43 ± 34.64	30.74 ± 31.83	49.18 ± 32.27
GP + RFTs	12.09 ± 42.79	4.22 ± 44.44	12.48 ± 39.91	47.32 ± 49.23
Lasso + Lasso	−60.94 ± 4.75	35.85 ± 60.29	−2.21 ± 62.41	22.51 ± 71.99
GP + Lasso	−17.89 ± 187.64	7.60 ± 88.89	−2.86 ± 72.57	21.89 ± 80.90
GP + MI	26.42 ± 38.7	−13.92 ± 56.41	6.94 ± 36.80	52.32 ± 46.38
GPARD + Full	18.22 ± 40.83	21.80 ± 41.17	13.30 ± 41.09	23.09 ± 36.14
GP + two fs	−0.00 ± 0.03	−2.04 ± 9.01	−0.02 ± 0.10	0.08 ± 0.74
Correlation coefficient				
RFTs + RFTs	0.57 ± 0.27	0.66 ± 0.28	0.57 ± 0.26	0.68 ± 0.27
GP + RFTs	0.46 ± 0.26	0.41 ± 0.27	0.45 ± 0.27	0.71 ± 0.25
Lasso + Lasso	0.35 ± 0.30	0.66 ± 0.29	0.41 ± 0.31	0.62 ± 0.29
GP + Lasso	0.41 ± 0.32	0.58 ± 0.28	0.45 ± 0.33	0.61 ± 0.31
GP + MI	0.56 ± 0.25	0.25 ± 0.22	0.41 ± 0.24	0.75 ± 0.22
GPARD + Full	0.47 ± 0.29	0.49 ± 0.31	0.43 ± 0.29	0.53 ± 0.23
GP + two fs	0.12 ± 0.43	0.21 ± 0.37	0.21 ± 0.30	0.36 ± 0.13

PDMS is polydimethylsiloxane (see Figure 1); RFT is random forest tree; GP is Gaussian Process, often used interchangeably with Gaussian Process Regression (GPR); LASSO is the least absolute shrinkage and selection operator; MI is mutual information (note that JMI, joint mutual information, has also been discussed in this study); GPARD is Gaussian Process Regression; GP + two fs is the Gaussian Process model run with two descriptors, log P and MW, to mimic the descriptors used in common QSAR models of skin permeation; Full refers to studies run with the full data set. Results highlighted in bold are those which achieved the best outcome for each test.

of ARD gives much better results. For example, all IONs of GPARD + Full are greater than 10%, whereas all ION values in Table 1 are around zero (for the GP method

using the full feature set). This is because, rather than using an identical length scale over all features and treating all features equally, GPARD uses a different

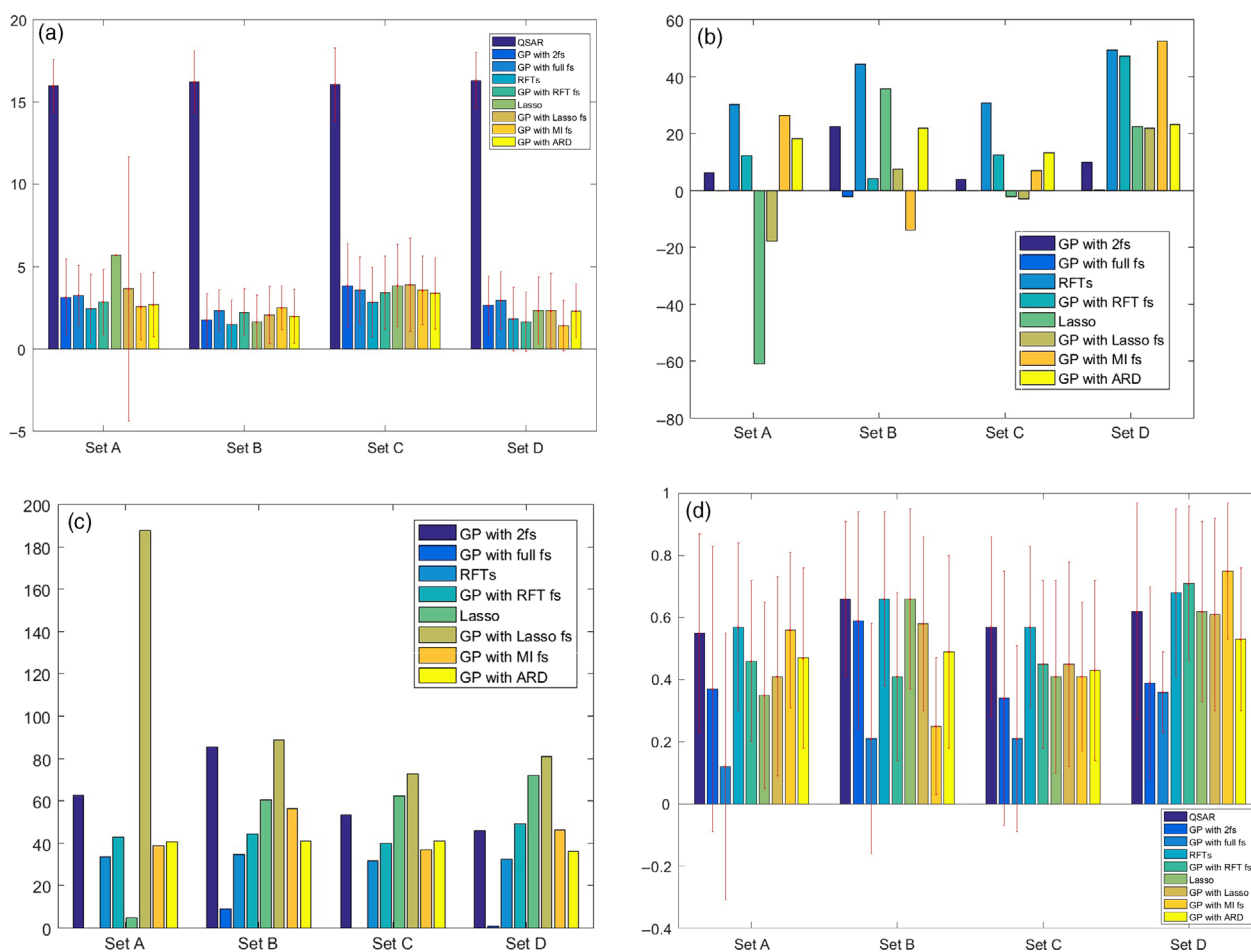


Figure 5 Overview of the outcome of statistical measures (MSE, ION and correlation coefficient) for the different QSAR and Machine Learning methods investigated. Results are shown as the mean and standard deviation following 10 000 repetitions of the validation set and presented as mean and standard deviation; (a) the grouped bar plot of mean values of MSE, with error bars also shown on the plot; (b) the mean values of ION, with the standard deviations of ION shown in (c) for clarity and (d) the grouped bar plot of mean values of correlation coefficients with error bars shown on the plot. [Colour figure can be viewed at wileyonlinelibrary.com]

length scale for each feature to indicate the importance of each feature in its predictive ability.

Figure 5a firstly shows that the benchmark QSAR models perform worst among all the models, which might be expected based on the above discussion and may reflect more widely the value of PDMS as an alternative to mammalian tissue.^[39] Figure 5b shows the mean values of ION, and it can be seen that there is no improvement if the GP model is used with the full data set compared to the naïve model. Interestingly, in Figure 5c the corresponding standard deviation is consistently low over all four validation sets, suggesting that the full feature data set contains a substantial amount of noise. Figure 5b also indicates that the performance of the LASSO models varies across the four validation sets; they are better than the naïve model for two data sets (A and C) and worse for the other two. RFTs give

better results than the naïve model for all four validation sets. The value obtained using GP with the full feature selection set is the lowest for all sets of data. The QSAR models yields relatively good correlation coefficients, which addresses the issues described above and also previously in attempting to demonstrate the potential of the PDMS membrane to represent human skin.^[39] Moreover, it indicates that use of the correlation coefficient alone – as is often the case in skin permeability studies – may not be a reliable indicator of model performance and that use and consideration of a series of statistical measures may provide more suitable outcomes in terms of describing model quality.

A small number of descriptors were repeatedly found in at least 5000 of the repeated experiments. The features selected using RFTs are shown in Table 3. From this

Table 3 Lists of physicochemical descriptors selected using RFTs method. The descriptors listed are those chosen more than 5000 times

Training set	Feature list	Software	Definition
TrnA	$\log P_{(o/w)}$	–	The logarithm of the octanol-water partition
	$\text{Sum}(E_a)/\text{Alpha}$	HYBOT	Hydrogen bonding descriptor – E_a factor value per unit of molecular volume
	CHssNH	MolconnZ	Molecular connectivity descriptor
	PEOE_VSA_FPOS	MOE	Partial charge descriptors – Fractional positive van der Waals surface area. This is the sum of the v_i such that q_i is non-negative divided by the total surface area. The v_i are calculated using a connection table approximation.
	SHC _{sat}	MolconnZ	E-State of Csp3 bonded to unsaturated C atoms
	GGI6	Dragon	Topological charge index of order 6
	Q_VSA_FPPOS	MOE	Partial charge descriptors – Fractional negative van der Waals surface area. This is the sum of the v_i such that q_i is negative divided by the total surface area. The v_i are calculated using a connection table approximation.
	$\text{Sum}(C_{ad})/\text{Alpha}$	HYBOT	Hydrogen bonding descriptor – C_a*d factor value per unit of molecular volume
	ALOGP	Dragon	Ghose-Crippen octanol-water partition coefficient
TrnB	$\log P_{(o/w)}$	–	
	CHssNH	MolconnZ	Molecular connectivity descriptor
	$\text{Sum}(C_{ad})/\text{Alpha}$	HYBOT	Hydrogen bonding descriptor – C_a*d factor value per unit of molecular volume
	BIC2	Dragon	Bond Information Content index (neighbourhood symmetry of 2-order)
	$\text{Sum}(E_a)/\text{Alpha}$	HYBOT	Hydrogen bonding descriptor – E_a factor value per unit of molecular volume
	Q_VSA_FPPOS	MOE	Partial charge descriptors – Fractional negative van der Waals surface area. This is the sum of the v_i such that q_i is negative divided by the total surface area. The v_i are calculated using a connection table approximation.
	ALOGP	Dragon	Ghose-Crippen octanol-water partition coefficient
	ALOGP2	Dragon	Squared Ghose-Crippen octanol-water partition coefficient, i.e. $(\log P)^2$
	NHB _{int2}	HYBOT	Atom type electrotopological state: Count of E-State descriptors of strength for potential Hydrogen Bonds of path length 2.
	$\text{Sum}(C_a)/\text{Alpha}$	HYBOT	Hydrogen bonding descriptor – C_a factor value per unit of molecular volume
	SIC2	Dragon	Information indices – Structural Information Content index (neighbourhood symmetry of 2-order)
	E _{sol}	MOE	Potential energy descriptor – Solvation energy. In the Potential Setup panel, the term enable flag is ignored, but the term weight is applied.
	PEOE_VSA_FPOS	MOE	Partial charge descriptor – Fractional positive van der Waals surface area. This is the sum of the v_i such that q_i is non-negative divided by the total surface area. The v_i are calculated using a connection table approximation.
	$\text{Sum}(E_{ad})/\text{Alpha}$	HYBOT	Hydrogen bonding descriptor – E_a*d factor value per unit of molecular volume
TrnC	CHssNH	MolconnZ	Molecular connectivity descriptor
	$\text{Sum}(E_a)/\text{Alpha}$	HYBOT	Hydrogen bonding descriptor – E_a factor value per unit of molecular volume
	PEOE_VSA_FPOS	MOE	Partial charge descriptor – Fractional positive van der Waals surface area. This is the sum of the v_i such that q_i is non-negative divided by the total surface area. The v_i are calculated using a connection table approximation.
	Q_VSA_FPPOS	MOE	Partial charge descriptors – Fractional negative van der Waals surface area. This is the sum of the v_i such that q_i is negative divided by the total surface area. The v_i are calculated using a connection table approximation.
	MATS1m	Dragon	2D autocorrelation – Moran autocorrelation of lag 1 weighted by mass
	SHC _{sat}	MolconnZ	E-State of Csp3 bonded to unsaturated C atoms
	GATS1m	Dragon	2D autocorrelation – Geary autocorrelation of lag 1 weighted by mass
	$\text{Sum}(C_{ad})/\text{Alpha}$	HYBOT	Hydrogen bonding descriptor – C_a*d factor value per unit of molecular volume
	CHssNH	MolconnZ	Molecular connectivity descriptor

analysis six common features have been identified from all four training sets, namely: $\log P_{(o/w)}$, PEOE_VSA_FPOS, Q_VSA_FPOS, CHssNH, $\text{Sum}(E_a)/\text{Alpha}$ and $\text{Sum}(C_{ad})$. Furthermore, between 28 and 39 features have been selected from each training set using MI. Seven common features were identified, namely b_{IrotR} , b_{rotR} , $\log P_{(o/w)}$, Q-

VSA_FPPOS, $v_{\text{surf_A}}$, $\text{Max}(E_a)*\text{Max}(E_d)$ and $\text{Sum}(E_a)/\text{Alpha}$. These descriptors are defined in Table 3 but, in general, they fall under three main categories in that they describe lipophilicity, partial charge and hydrogen bonding as key determinants of PDMS permeation. This is consistent with many previous findings in this field.^[59,63,64] but

Table 3 (Continued)

Training set	Feature list	Software	Definition
TrnD	$\log P_{(o/w)}$	–	
	Sum(E_a)/Alpha	HYBOT	Hydrogen bonding descriptor – E_a factor value per unit of molecular volume
	PEOE_VSA_FPOS	MOE	Partial charge descriptor – Fractional positive van der Waals surface area. This is the sum of the v_i such that q_i is non-negative divided by the total surface area. The v_i are calculated using a connection table approximation.
	Sum(C_{ad})/Alpha	HYBOT	Hydrogen bonding descriptor – C_{ad} factor value per unit of molecular volume
	Sum(C_a)/Alpha	HYBOT	Hydrogen bonding descriptor – C_a factor value per unit of molecular volume
	Q_VSA_FPPOS	MOE	Partial charge descriptors – Fractional negative van der Waals surface area. This is the sum of the v_i such that q_i is negative divided by the total surface area. The v_i are calculated using a connection table approximation.
	SHC _{sat}	MolconnZ	E-State of Csp3 bonded to unsaturated C atoms
	E _{sol}	MOE	Potential energy descriptor – Solvation energy. In the Potential Setup panel, the term enable flag is ignored, but the term weight is applied.
	Lip _{acc}	MOE	The number of O and N atoms
	GCUT_SLOGP_1	MOE	Adjacency and Distance Matrix Descriptors – The GCUT descriptors using atomic contribution to logP (using the Wildman and Crippen SlogP method) instead of partial charge
	ALOGP	Dragon	Ghose-Crippen octanol-water partition coeff. (logP)
	ALOGP2	Dragon	Squared Ghose-Crippen octanol-water partition coeff. (logP) ²
	BLTA96	Dragon	Verhaar Algae base-line toxicity from MLOGP (mmol/l)
	NHB _{int} 2	HYBOT	Atom type electrotopological state: Count of E-State descriptors of strength for potential Hydrogen Bonds of path length 2.
	BIC2	Dragon	Bond Information Content index (neighbourhood symmetry of 2-order)

provides a more specific molecular basis in determining specific structural features that contribute positively or negatively to skin permeation.

Figure 6 indicates that a number of the significant descriptors demonstrate a reasonable correlation. In Figure 7, it can be seen that, using the MI method, b_{1rotR} (the fraction of rotatable single bonds) and b_{rotR} (the fraction of rotatable bonds) are highly correlated. The results of four independent test sets with different descriptor inputs, obtained using the RFT and MI methods (Table 4) predominately suggest that the best results all come from the selected descriptors (i.e. using either RFTs or MI). These results suggest that the two best predictive models are those which are produced using the feature selection method to generate the most relevant descriptors for the data set analysed. The consideration of highly correlated descriptors (Table 5) indicates that statistically improved models are generated when highly correlated descriptors, for example b_{1rotR} and b_{rotR} , are removed from the model (that is, only one or the other, not both, are included in the final model). This highlights the need to consider the underlying statistical nature of the data and the potential pitfalls of using highly correlated or covariate descriptors in models simply because the data and particular descriptors are easily generated or readily available.^[30,32,36] The consideration of highly correlated descriptors is shown in Table 5, where the results displayed

are obtained from a MI analysis with and without highly correlated descriptors. The top line of Table 5 shows data obtained using six common descriptors, which indicates that this group of descriptors gives slightly better results over all test sets. This indicates that the 6-descriptor model outperforms the 8-descriptor model, yielding a significant improvement on test sets A and D.

The modelling approach in this study is generally perceived to be of a ‘top-down’ nature, in that it provides a gross estimate of skin permeability and, like related QSAR-type methods, offers limited mechanistic insight into the permeation process.^[36] However, the addition of feature selection methodologies to this problem domain in order to precisely determine the molecular descriptors of interest offers an opportunity to expand this approach to generate more precise information about the membrane permeation process. However, a practical limitation to this approach is the size of the data set ($n = 77$). For example many previous studies, mostly in the 1990s, examined subsets of skin permeation generated from Flynn’s data set.^[33,65–71] In many cases, these subsets were focused on specific molecular properties (such as non-electrolytes^[71]) or specific functional groups.^[66,67] One perceived issue with these studies is that they resulted in poor statistical fits in many cases, which one might relate to the small size of the data sets used or, for example, the

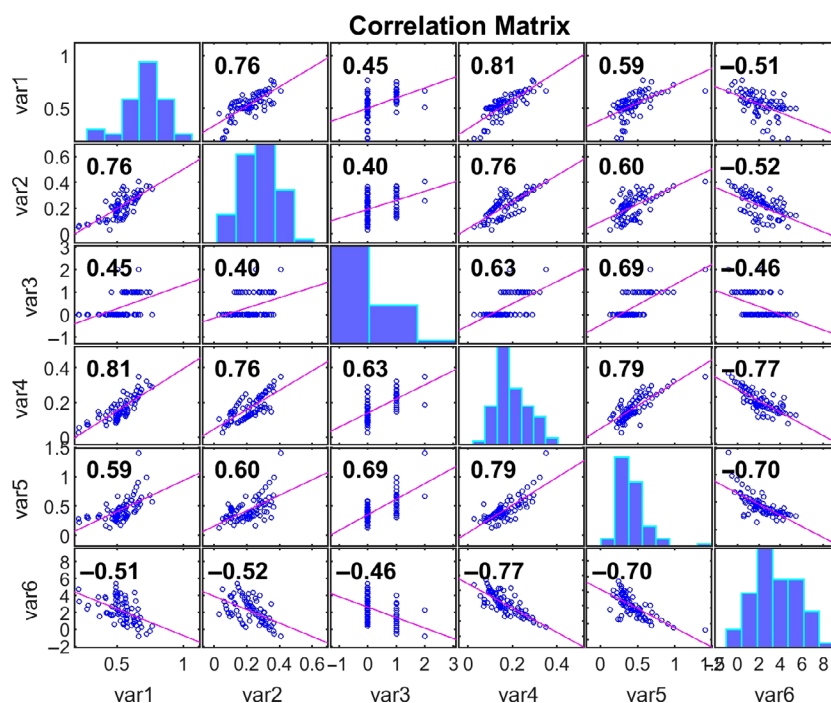


Figure 6 The correlation coefficient matrix plot for common features selected using RFTs, where *var1* to *var6* denote, respectively, $\log P_{o/w}$, PEOE_VSA_FPOS, Q_VSA_FPPOS, CHssNH, Sum(E_a)/Alpha and Sum(C_{ad})/Alpha, respectively. [Colour figure can be viewed at wileyonlinelibrary.com]

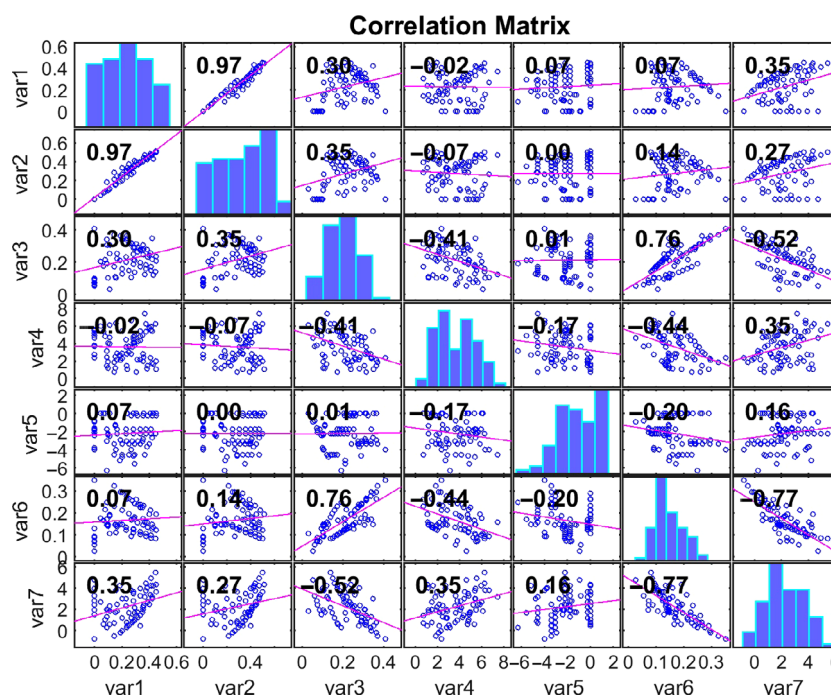


Figure 7 The correlation coefficient matrix plot for common features selected using MI, where *var1* to *var6* denote, respectively, b_{1rotR} , b_{rotR} , $\log P_{o/w}$, Q_VSA_FPPOS, vsurf_A, Max(E_a)*Max(E_d) and Sum(E_a)/Alpha, respectively. [Colour figure can be viewed at wileyonlinelibrary.com]

Table 4 Results for MSE, ION and correlation coefficient for the independent test sets

Test and/or descriptors used	Test Set A	Test Set B	Test Set C	Test Set D
MSE				
QSAR	19.07	15.64	16.60	14.09
2fs	6.43	16.60	0.52	11.07
Full	7.01	16.11	1.30	10.45
RFT _{common}	5.28	14.83	0.23	12.32
MI _{common}	4.68	15.36	0.40	8.70
ION				
QSAR	−150.47	2.89	−0.00	−34.79
2fs	15.55	−3.04	60.43	−5.85
Full	−0.00	0.00	0.00	−0.00
RFT _{common}	30.70	7.92	82.18	−17.83
MI _{common}	38.50	4.62	69.42	16.78
Correlation coefficient				
QSAR	0.33	0.61	0.54	0.21
2fs	0.50	0.68	0.80	−0.28
Full	NaN	0.32	0.06	−0.48
RFT _{common}	0.84	0.57	0.92	−0.18
MI _{common}	0.85	0.58	0.83	0.53

NaN indicates that, when using the full set of descriptors, all predictions have a fixed value for this test set, and therefore, the correlation coefficients cannot be computed; other abbreviations in this table are as defined in the footnotes of Tables 1 and 2. Results highlighted in bold are those which achieved the best outcome for each test.

Table 5 Results for ION obtained from RFTs and MI analyses for the independent test sets with and without highly correlated features

Test and number of descriptors used	Test Set A	Test Set B	Test Set C	Test Set D
RFT – 6 descriptors	30.70	7.92	82.18	−17.83
RFT – 7 descriptors	26.42	5.69	76.43	−24.21
MI – 6 descriptors	38.50	4.62	69.42	16.78
MI – 8 descriptors	27.84	1.27	67.64	6.29

Abbreviations in this table are as defined in the footnotes of Tables 1 and 2. Results highlighted in bold are those which achieved the best outcome for each test.

absence of methods that can optimise outputs from small datasets^[32]. Thus, while it is possible to speculatively identify specific molecular fragments that would improve permeation models, validation of such models might be difficult and impractical to develop due to the size of the data sets involved.

This study raises a novel and highly significant finding in the field of modelling in percutaneous absorption. Modelling of skin permeation has been shaped historically by models that used as few as two descriptors and assumptions were subsequently made about the mechanism of skin permeation that shaped detrimentally subsequent developments in this field. It is known that descriptors such as melting point,

ionisation and hydrogen bonding play a role in skin permeation. The main point, therefore, of this study was to propose a methodology that examines as many descriptors as possible in an unbiased manner and which then selects the most appropriate descriptors based on established methods, thus providing a platform for the development of models to predict permeation across, in this case, PDMS membranes. While such analogies cannot be automatically extended to fields other than percutaneous absorption, it should be noted that similar methods have been applied to a range of biological and environmental endpoints. Thus, the findings of this study clearly have a wider relevance beyond percutaneous absorption.

Conclusions

Two sets of molecular descriptors which can provide improved predictions, compared to the use of either two descriptors (log P and MW) or a full list of excess of two thousand descriptors, of permeability across a PDMS membrane, have been identified (Table 3). Using the GP method with either of these two sets of descriptors provides significantly better predictions of permeability than using a QSAR model, when compared with four independent test sets, although the use of such benchmarks in this study has clear limitations. The generation of two sets of descriptors echoes previous findings^[30] that suggest certain permutations of different descriptors yield similarly relevant predictive models. More broadly in the development of reliable, robust and valid models it is also recommended that highly correlated descriptors are removed from the data set when using the GP method. It is also apparent that re-sampling sub-training sets helps with feature selection for small data sets, and it is again recommended that this approach be taken in the development of small data sets.

The implications of these results for building relevant predictive models are novel and significant. Models in the field of percutaneous absorption are based on a small number of significant descriptors which have been widely used and often assumed to be relevant based on historical experimental findings. While it has been shown in this study that some of these descriptors are not relevant, they are still employed routinely in model development. This study indicates that this practice should be discontinued and that future predictive models will be improved by the analysis adopted in this study of the descriptors employed, rather than simply using any and all available descriptors or using a smaller set of descriptors whose choice has been influenced empirically by previously published *in vitro* diffusion studies.

Declarations

Conflicts of interest

The Authors have no conflicts of interest to report.

Acknowledgements and funding

This research did not receive any specific grant from funding agencies in the public, commercial or not-for-profit sectors. The authors would like to thank the University of

Hertfordshire and Keele University for supporting this study and Professor Mark Cronin (Liverpool John Moores University) for assistance with the data set.

Author contribution

This study was designed and led by GM and YS. Data sets and descriptors were developed and generated by DG and MH. GP analysis was carried out by RA, ND and YS. Analysis of the data outputs and interpretation and contextualisation for skin permeation were written by GM and SW.

References

- Petitot F *et al.* Incorporation and distribution of uranium in rats after a contamination on intact or wounded skin. *Health Phys* 2007; 5: 464–474.
- Vallet V *et al.* In vitro percutaneous penetration of organophosphorus compounds using full-thickness and split-thickness pig and human skin. *Toxicol In Vitro* 2007; 6: 1182–1190.
- Caussin J *et al.* Lipid organization in human and porcine stratum corneum differs widely, while lipid mixtures with porcine ceramides model human stratum corneum lipid organization very closely. *Biochim Biophys Acta* 2008; 6: 1472–1482.
- Klang V *et al.* In vitro vs. in vivo tape stripping: validation of the porcine ear model and penetration assessment of novel sucrose stearate emulsions. *Eur J Pharm Biopharm* 2012; 3: 604–614.
- Woolfson AD *et al.* Development and characterisation of a moisture-activated bioadhesive drug delivery system for percutaneous local anaesthesia. *Int J Pharm* 1998; 169: 83–94.
- Khan GM *et al.* Assessment of drug permeability distributions in two different model skins. *Int J Pharm* 2005; 303: 81–87.
- Wasdo SC *et al.* A correlation of flux through a silicone membrane with flux through hairless mouse skin and human skin in vitro. *Int J Pharm* 2009; 373: 62067.
- Gullick DR *et al.* Formulation and characterization of a captopril ethyl ester drug-in-adhesive-type patch for percutaneous absorption. *Drug Dev Ind Pharm* 2010; 36: 926–932.
- Watkinson AC *et al.* Is it logical to reject skin samples based on water permeability data? In: Brain KR, Walters KA, eds. *Perspectives in Percutaneous Penetration*, vol. 6A. Cardiff, UK: STS Publishing, 1998: 62.
- Pellet MA *et al.* Diffusion of deuterated ethanol in synthetic membranes using ATR-FTIR spectroscopy. *Eur J Pharm Sci* 1994; 2: 155.
- Ley EE, Bunge AL. Chemical transport in silicone rubber membranes from pure powders and saturated aqueous solutions. *J Memb Sci* 2007; 292: 35–44.
- Dias M *et al.* Influence of membrane-solute interactions on solute permeation in model systems. *Int J Pharm* 2007; 336: 108–114.
- McAuley WJ *et al.* Simultaneous monitoring of drug and solvent diffusion across a model membrane using ATR-FTIR spectroscopy. *Eur J Pharm Sci* 2009; 38: 378–383.
- McAuley WJ *et al.* Thermodynamic considerations of solvent/enhancer uptake into a model membrane. *Int J Pharm* 2010; 396: 134–139.
- Frum Y *et al.* Evidence that drug flux across synthetic membranes is described by normally distributed permeability coefficients. *Eur J Pharm Biopharm* 2007; 67: 434–439.
- Liu P *et al.* Variation of in vitro skin permeation data for ionic compounds. *Pharm Res* 1991; 8: S137.
- Williams AC *et al.* On the non-Gaussian distribution of human skin permeabilities. *Int J Pharm* 1992; 86: 69–77.
- Cornwell PA, Barry BW. Effects of penetration enhancer treatment on the statistical distribution of human skin permeabilities. *Int J Pharm* 1995; 117: 101–112.
- Roper CS *et al.* Tritiated water permeability coefficient assessment and rejection criteria for barrier function of human skin. In: Brain KR, Walters KA, eds. *Perspectives in Percutaneous Penetration*, vol. 7A. Cardiff, UK: STS Publishing, 2000.
- Fasano WJ *et al.* Rapid integrity assessment of rat and human epidermal membranes for in vitro dermal regulatory testing: correlation of electrical resistance with tritiated water permeability. *Toxicol In Vitro* 2002; 16: 731–740.
- Wenkers BP, Lippold BC. The statistical distribution of in vivo skin permeability. *Pharm Pharmacol Commun* 1999; 5: 425–427.
- Chen Y *et al.* Prediction of flux through polydimethylsiloxane membranes using atomic charge calculations. *Int J Pharm* 1993; 94: 81–88.
- Chen Y *et al.* Prediction of flux through polydimethylsiloxane membranes using atomic charge calculations: application to an extended data set. *Int J Pharm* 1996; 137: 149–158.
- Cronin MTD *et al.* An investigation of the mechanism of flux across polydimethylsiloxane membranes by the use of quantitative structure-permeability relationships. *J Pharm Pharmacol* 1998; 50: 143–152.

25. Agatonovic-Kustrin S *et al.* ANN modelling of the penetration across a polydimethylsiloxane membrane from theoretically derived molecular descriptors. *J Pharm Biomed Anal* 2001; 26: 241–254.
26. Geinoz S *et al.* Quantitative structure-permeation relationships for solute transport across silicone membranes. *Pharm Res* 2002; 19: 1622–1629.
27. Moss GP *et al.* Design, synthesis and characterisation of captopril prodrugs for enhanced percutaneous absorption. *J Pharm Pharmacol* 2006; 58: 167–177.
28. Sun Y *et al.* The application of Gaussian Processes in the predictions of permeability across mammalian membranes. In: Vila A, Duch W, Masulli F, Palm G, eds. *Artificial Neural Networks and Machine Learning, ICANN 2012*, Springer, 2012: 507–514.
29. Sun Y *et al.* The application of Gaussian Processes in the prediction of permeability across mammalian and polydimethylsiloxane membranes. *Artif Intell Res* 2012; 1: 86–98.
30. Lam LT *et al.* The application of feature selection to the development of Gaussian Process models for percutaneous absorption. *J Pharm Pharmacol* 2010; 62: 738–749.
31. Sun Y *et al.* The application of stochastic machine learning methods in the prediction of skin penetration. *App Soft Comp* 2011; 11: 2367–2375.
32. Ashrafi P *et al.* Model fitting for small skin permeability data sets: hyperparameter optimisation in Gaussian Process Regression. *J Pharm Pharmacol* 2018; 70: 361–373.
33. Flynn GL. Physicochemical determinants of skin absorption. In: Gerrity TR, Henry CJ, eds. *Principles of Route-to-Route Extrapolation for Risk Assessment*. New York, NY: Elsevier, 1992: 93–127.
34. Johnson ME *et al.* Permeation of steroids through human skin. *J Pharm Sci* 1995; 84: 1144–1146.
35. Moss GP, Cronin MTD. Quantitative structure-permeability relationships for percutaneous absorption: re-analysis of steroid data. *Int J Pharm* 2002; 238: 105–109.
36. Moss GP *et al.* *Predictive Methods in Percutaneous Absorption*. Heidelberg: Springer AG, 91–102. <https://www.springer.com/gb/book/9783662473702>
37. Moss GP *et al.* The application of Gaussian Processes to the prediction of percutaneous absorption. *J Pharm Pharmacol* 2009; 61: 1147–1153.
38. Moss GP *et al.* Quantitative structure-permeability relationships (QSPRs) for percutaneous absorption. *Tox In Vitro* 2002; 16: 299–317.
39. Moss GP *et al.* The application and limitations of mathematical modelling in the prediction of permeability across mammalian skin and polydimethylsiloxane membranes. *J Pharm Pharmacol* 2011; 63: 1411–1427.
40. van der Maaten L, Hinton GE. Visualizing high-dimensional data using t-sne. *J Mach Learn Res* 2008; 9: 2579–2605.
41. Fedors RF. A method for estimating both the solubility parameters and molar volumes of liquids. *Polymer Eng Sci* 1974; 14: 147–154.
42. Saeyns Y *et al.* A review of feature selection techniques in bioinformatics. *Bioinformatics* 2007; 23: 2507–2517.
43. Efromovick S. Adaptive estimation of error density in nonparametric regression with small sample size. *J Stat Plan Interference* 2007; 137: 363–378.
44. Sima C, Dougherty ER. What should be expected from feature selection in small-sample settings. *Bioinformatics* 2006; 22: 2430–2436.
45. Braga-Neto U, Dougherty E. Bolstered error estimation. *Pattern Recog* 2004; 37: 1267–1281.
46. Bornelov S, Komorowski J. Selection of significant feature using Monte Carlo feature selection. In: Matwin S, Mielniczuk J eds. *Challenges in Computational Statistics and Data Mining, Studies in Computational Intelligence*, Vol. 605. Switzerland: Springer International Publishing, 2016: 25–38.
47. DuPont WD, Plummer WD. Power and sample size calculations for studies involving linear regression. *Control Clin Trials* 1998; 19: 589–601.
48. Golugula A *et al.* Evaluating feature selection strategies for high dimensional, small sample size datasets. In: *33rd Annual International Conference of the IEEE EMBS*, IEE EMBS, Boston, Massachusetts, USA, 2011. <http://engineering.case.edu/centers/ccipd/sites/ccipd.case.edu/files/publications/Evaluating-feature-selection-strategies-for-high-dimensional-small-sample-size-datasets.pdf>
49. Bermingham ML *et al.* Application of high-dimensional feature selection: evaluation for genomic prediction in man. *Sci Rep* 2015; 5: Article Number 10312.
50. Li L, Yao W. Fully Bayesian logistic regression with hyper-LASSO priors for high-dimensional feature selection. *J Stat Comp Simul* 2018; 88: 2827–2851.
51. Ho TK. Random decision forests. In: *Proceedings of the 3rd International Conference on Document Analysis and Recognition*, Montreal, Canada, 1995: 278–282.
52. Breiman L. Random forests. *Mach Learn* 2001; 45: 5–32.
53. Loh W-Y. Regression trees with unbiased variable selection and interaction detection. *Statistica Sinica* 2002; 12: 361–386.
54. Tibshirani R. Regression shrinkage and selection via the lasso. *J R Stat Soc Ser B* 1996; 58: 267–288.
55. Yang H, Moody J. Data visualisation and feature selection: new algorithms for non-Gaussian data. In: *Advances in Neural Information Processing Systems 12*. Cambridge: The MIT Press, 2000: 687–693. <https://papers.nips.cc/paper/1779-data-visualization-and-feature-selection-new-algorithms-for-non-gaussian-data.pdf>
56. Brown G *et al.* Conditional likelihood maximisation: a unifying framework for information theoretic feature selection. *J Mach Learn Res* 2012; 13: 27–66.
57. Rasmussen CE, Williams CKI. *Gaussian Processes for Machine Learning*. Cambridge: The MIT Press, 2006.
58. Radford NM. *Bayesian Learning for Neural Networks*. Heidelberg: Springer AG, 1996. <https://www.springer.com/gb/book/9780387947242>

59. Roberts MS *et al.* Epidermal permeability – penetrant structure relationships. 1. An analysis of methods of predicting penetration of monofunctional solutes from aqueous solutions. *Int J Pharm* 1995; 126: 219–233.
60. Potts RO, Guy RH. Predicting skin permeability. *Pharm Res* 1992; 9: 663–669.
61. Moss GP *et al.* The application and limitations of mathematical models in the prediction of permeability across mammalian skin and polydimethylsiloxane membranes. *J Pharm Pharmacol* 2011; 63: 1411–1427.
62. Ashrafi P *et al.* The influence of diffusion cell type and experimental temperature on machine learning models of skin permeability. *J Pharm Pharmacol* 2020; 72: 197–208.
63. Roberts MS *et al.* Epidermal permeability – penetrant structure relationships. 2. The effect of H-bonding groups in penetrants on their diffusion through the *stratum corneum*. *Int J Pharm* 1996; 132: 23–32.
64. Pugh WJ *et al.* Epidermal permeability – penetrant structure relationships. 3. The effect of hydrogen bonding interactions and molecular size on diffusion across the *stratum corneum*. *Int J Pharm* 1996; 138: 149–165.
65. Abraham MH *et al.* Hydrogen bonding part 46. A review of the correlation and prediction of transport properties by an LFER method: physicochemical properties, brain penetration and skin permeability. *Pest Sci* 1999; 55: 78–88.
66. Abraham MH *et al.* The factors that influence skin penetration of solutes. *J Pharm Pharmacol* 1995; 47: 8–16.
67. Abraham MH *et al.* Algorithms for skin permeability using hydrogen bond descriptors: the problem of steroids. *J Pharm Pharmacol* 1997; 49: 858–865.
68. Abraham MH *et al.* Hydrogen bonding part 16. A new solute solvation parameter π -h2 from gas chromatographic data. *J Chromatogr* 1991; 587: 213–228.
69. Abraham MH. Scales of hydrogen bonding: their construction and application to physicochemical and biochemical processes. *Chem Soc Rev* 1993; 22: 73–83.
70. Barratt MD. Quantitative structure–activity relationships for skin permeability. *Tox In Vitro*. 1995; 9: 27–37.
71. Potts RO, Guy RH. A predictive algorithm for skin permeability: the effects of molecular size and hydrogen bond activity. *Pharm Res* 1995; 12: 1628–1633.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Appendix S1. Methodology for the study.

Appendix S2. PDMS descriptors.